

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Converting pH 1:1 H₂O and 1:2_{CaCl₂} to 1:5 H₂O to contribute to a harmonized global soil database[☆]



Zamir Libohova^{a,*}, Skye Wills^a, Nathan P. Odgers^b, Richard Ferguson^a, Rick Nesser^a, James A. Thompson^c, Larry T. West^a, Jonathan W. Hempel^a

^a U.S. Department of Agriculture-Natural Resources Conservation Service, 100 Centennial Mall North, Lincoln, NE 68508, USA

^b The University of Sydney, Faculty of Agriculture and Environment, Biomedical Building C81, University of Sydney, New South Wales 2006, Australia

^c West Virginia University, Division of Plant and Soil Sciences, PO Box 6108, Morgantown, WV 26506-6108, USA

ARTICLE INFO

Article history:

Received 28 September 2012

Received in revised form 13 August 2013

Accepted 16 August 2013

Available online 5 October 2013

Keywords:

Soil pH

Pedotransfer functions

GlobalSoilMap

ABSTRACT

The GlobalSoilMap initiative calls for the generation of continuous maps for soil properties, including pH in a 1:5 suspension of soil in water (pH 1:5_w) based on a standard method, ISO 10390. The United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS) employs a 1:1 suspension of soil in water (pH 1:1_w), and a 1:2 suspension of soil in _{CaCl₂} (0.01 M) (pH 1:2_{CaCl₂}) for routine pH analysis (Soil Survey Staff, 2009). The objective of this study was to determine the most efficient way to convert these pH values to the GlobalSoilMap standard. For this analysis, 563 soil samples from the USDA-NRCS-National Soil Survey Center (NSSC) soil archive, which had been previously analysed for pH 1:1_w and pH 1:2_{CaCl₂}, were selected for determination of pH 1:5_w, pH 1:5_{CaCl₂} and electrical conductivity (EC) in 1:2 suspension of soil in water (EC 1:2_w). The samples represented 11 soil orders, 8 mineralogy classes, 5 family particle size classes, 4 genetic master horizons, and 7 depth intervals. For each category, 25–30 samples were selected to represent a comprehensive pH range. Regression analysis showed strong and significant relationships ($R^2 > 0.92$) between pH methods across all categories. The simple linear regression equation, $\text{pH } 1:5_w = -0.51 + 1.06 \text{ pH } 1:1_w$, had an RMSE = 0.44 pH units. Smoothing spline, did not significantly improve pH 1:5_w predictions, nor did the incorporation of EC. Genetic horizons and soil depth intervals did not have a significant effect on pH 1:5_w. The linear regression models for predicting pH 1:5_w using pH 1:1_w or pH 1:2_{CaCl₂} as predictors emerged as the best candidates for a standard pedotransfer function. Using pedotransfer functions such as these will allow for the simple conversion of existing measured and estimated pH 1:1_w or pH 1:2_{CaCl₂} values from NRCS databases to the GlobalSoilMap standard of pH 1:5_w.

Published by Elsevier B.V.

1. Introduction

Soil pH, which is an important measure of soil acidity, nutrient availability, and soil productivity, is considered a master variable (Brady and Weil, 1999). Soil pH influences many biogeochemical processes in the soil. Furthermore, pH is measured during typical analysis of soil samples, and is indirectly used as a key property in soil classification systems (Soil Survey Staff, 1999). The USDA-NRCS Kellogg Soil Science Laboratory (KSSL) routinely determines pH 1:1_w and/or pH 1:2_{CaCl₂} (Miller and Kissel, 2010; Soil Survey Staff, 2009). As of 2012, 300,000 samples from more than 50,000 pedons from across the United States and around the world have been analysed. This data is available through various portals such as Web Soil Survey (USDA-NRCS, 2011). These two methods are among many different laboratory procedures used to assess soil pH in

soils around the world, with methods varying in the ratio of soil to solution (e.g., 1:1, 1:2, 1:5) and the salt used in the solution (e.g., none, _{CaCl₂}, KCl). The GlobalSoilMap initiative calls for the generation of continuous maps of soil pH 1:5_w (pH measured for a ratio of 1 part soil and 5 parts water solutions) for the entire world for standard depth increments (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm) (GlobalSoilMap, 2012).

In the absence of a single, universally-used method for determining soil pH, the GlobalSoilMap specifications refer to the soil science literature for functions to convert between the various pH methods (GlobalSoilMap, 2012), including the work of Conyers and Davey (1988), Bruce et al. (1989), Aitken and Moody (1991), and Miller and Kissel (2010) (Table 1). In most of this work, the relationships developed for converting between methods were based on relatively small or geographically-limited sample sets. For example, Bruce et al. (1989) used 182 soil samples and Aitken and Moody (1991) used 90 samples to develop relationships for converting between pH 1:5_{CaCl₂} and pH 1:5_w for acidic soils in Australia. Conyers and Davey (1988) developed relationships for converting between pH 1:5_{CaCl₂} and pH 1:5_w

[☆] For inquiring about the data used in this article, please check webmaster@lin.usda.gov.

* Corresponding author.

Table 1
Models for relating various pH methods.

No. obs.	Predicted	Predictor(s)	Model	R ²	Source
11583	pH 1:5Ca ^a	pH 1:5 _W , EC 1:5 _W	Linear	0.94	Minasny et al. (2011)
11583	pH 1:5Ca	pH 1:5 _W , EC 1:5 _W	Curvilinear	0.95	Minasny et al. (2011)
120	pH 1:1 _W	pH 1:1Ca ^b , EC 1:1 _W	Exponential	0.99	Miller and Kissel (2010)
120	pH 1:2 _W	pH 1:1Ca ^b , EC 1:1 _W	Exponential	0.98	Miller and Kissel (2010)
70465	pH 1:5Ca	pH 1:5 _W	Smoothing spline	0.96	Henderson and Bui (2002)
7894	pH 1:5Ca	pH 1:5 _W	Linear	0.93	Ahern et al. (1995)
7894	pH 1:5Ca	pH 1:5 _W	Second order polynomial	0.93	Ahern et al. (1995)
7894	pH 1:5Ca	pH 1:5 _W	Third order polynomial	0.94	Ahern et al. (1995)
1342	pH 1:5Ca	pH 1:5 _W	Linear	0.89	Little (1992)
1342	pH 1:5Ca	pH 1:5 _W	Second order polynomial	Na	Little (1992)
1342	pH 1:5Ca	pH 1:5 _W	Third order polynomial	Na	Little (1992)
90	pH 1:5Ca	pH 1:5 _W	Linear	0.88	Aitken and Moody (1991)
90	pH 1:5Ca	pH 1:5 _W	Second order polynomial	0.92	Aitken and Moody (1991)
90	pH 1:5Ca	pH 1:5 _W	Linear	0.90	Aitken and Moody (1991)
90	pH 1:5Ca	pH 1:5 _W	Second order polynomial	0.94	Aitken and Moody (1991)
182	pH 1:5Ca	pH 1:5 _W	Linear	0.68	Bruce et al. (1989)
11	pH 1:5Ca	pH 1:5 _W	Linear	0.99	Conyers and Davey (1988)

^a pH (1:5 0.01 M CaCl₂).

^b pH (1:1 0.01 M CaCl₂).

based on 11 soil samples from Australia. Miller and Kissel (2010) analysed 120 soil samples from 33 US states and two Canadian provinces and developed relationships for various methods including pH 1:1_W, pH 1:1CaCl₂, pH 1:2_W, pH 1:2CaCl₂, and pH of saturated paste extract. Large datasets were used for two studies relating pH 1:5_W and 1:5CaCl₂. Ahern et al. (1995) developed relationships based on 7894 soil samples and Henderson and Bui (2002) based on 70,465 soil samples both from Australia. We know of no studies that have developed relationships for converting from pH 1:1_W and 1:2CaCl₂ to pH 1:5_W including those cited by GlobalSoilMap. While a few studies have investigated the influence of factors such as soil depth on pH (Bruce et al., 1989) the authors found none that evaluate the relationship of soil order, texture and mineralogy to soil pH.

Many researchers have found that the relationship between pH 1:5_W and pH 1:5CaCl₂ is not linear across the entire range of measured soil pH values (Ahern et al., 1995; Aitken and Moody, 1991; Bruce et al., 1989; Henderson and Bui, 2002). Various mechanisms have been offered in soil science literature to explain the non-linearity, such as: (i) calcium carbonate buffering for high soil pH; (ii) displacement and hydrolysis of Al³⁺ from exchange sites by CaCl₂ for low soil pH (Little, 1992), and (iii) variable charge theory (Aitken and Moody, 1991). This work has resulted in the development of various calibration equations from linear functions to higher-order polynomials and spline functions (Ahern et al., 1995; Aitken and Moody, 1991; Bruce et al., 1989; Conyers and Davey, 1988; Henderson and Bui, 2002; Little, 1992; Miller and Kissel, 2010; Minasny et al., 2011). When comparing pH 1:5_W and pH 1:5CaCl₂ methods, Aitken and Moody (1991) found a quadratic function to fit the data better than a simple linear regression, but Ahern et al. (1995) found that the quadratic function was not suitable for a wide range of pH values. Little (1992) found a cubic function that best described the relationships for pH's <5 and >7. Henderson and Bui (2002) found the smoothing spline function with six degrees of freedom to be a better fit for a wide range of pH values from 2.5 to 10.5. More recently, Minasny et al. (2011) included electrical conductivity (EC) in the prediction of pH 1:5CaCl₂ from 1:5_W, which improved predictions for acid and alkaline pH values.

The National Cooperative Soil Characterization Database (SCDB) contains data on more than 50,000 soil samples (National Cooperative Soil Survey, 2011). This extensive database affords an excellent opportunity to establish relationships between methods of soil pH measurement using soil samples that represent many soil orders, mineralogy and family particle size classes, as well as different genetic horizons and soil depths, which have not been the case for the other studies.

The objectives of this study were to (i) develop statistical relationships between soil pH 1:1_W, 1:2CaCl₂, EC and 1:5_W using a variety of methods,

and to (ii) discern the influence, if any, of other categorical soil variables on those relationships. The development of these relationships will enable the conversion of soil pH 1:1_W and pH 1:2CaCl₂ as routinely determined by the National Soil Science Center-Kellogg Soil Science Laboratory (NCSS-KSSL) to the soil pH 1:5_W required to meet GlobalSoilMap specifications. The selected pedotransfer functions will be used to convert the State Soil Geographic STATSGO2 (Soil Survey Staff, 2011a,b) database pH values to pH 1:5_W that will then be used to generate predictive soil pH maps that conform to the GlobalSoilMap specifications similar to the ones produced by Odgers et al. (2012) for soil organic carbon.

2. Material and methods

2.1. Soil samples

A total of 563 soil samples representing 98 pedons from the SCDB and physically archived at the KSSL that had been previously analysed for soil pH 1:1_W and pH 1:2CaCl₂ were selected for determination of soil pH 1:5_W, pH 1:5CaCl₂, and EC. In this paper, only the results for the conversion of soil pH 1:1_W and pH 1:2CaCl₂ to pH 1:5_W will be presented along with the influence of EC. The soil samples represented 11 soil orders (Alfisols, Andisols, Aridisols, Entisols, Gelisols, Histosols, Inceptisols, Mollisols, Spodosols, Ultisols, and Vertisols; Soil Survey Staff, 1999); 8 mineralogy classes (Amorphic, Glassy, Glassy over amorphic, Halloysitic, Isotonic, Kaolinitic, Mixed, and Smectitic; Soil Survey Staff, 2010); 5 family particle size classes (Clayey, Coarse loamy, Fine loamy, Fine silty, and Sandy; Schoeneberger et al., 2012), 4 genetic master soil horizons (O, A, B, and C) (Soil Survey Staff, 1999); and 7 depth intervals (0–10 cm, 10–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, 100–200 cm, and 200–350 cm). Each class had 25 to 30 samples that represented a comprehensive range of soil pH values from 2.9 to 10.5 and a wide geographic distribution (Fig. 1). The spatial distribution of point locations shows uneven representation of the southwestern US because the deserts have not been sampled as intensively as the agricultural areas. However, an adequate sample size of 24 samples from Aridisols represents soils of high importance for agriculture.

2.2. Analytical procedures

All soil samples had been previously analysed for pH 1:1_W and 1:20.01 M CaCl₂ according to the methods used by the KSSL (Soil Survey Staff, 2009). It was assumed that the likelihood of major changes in the pH is very low for dried samples that are stored in closed containers and at room temperatures. pH 1:5_W and pH 1:5CaCl₂ were determined

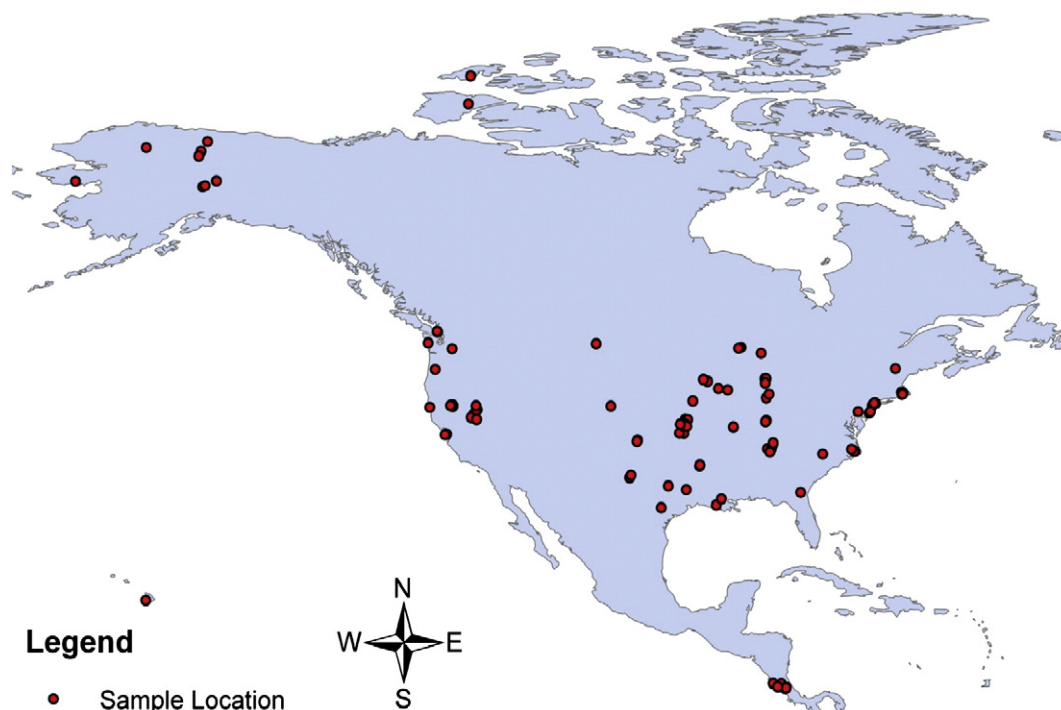


Fig. 1. The spatial distribution of sample locations for North America.

based on the methods described by Rayment and Higginson (1992). Electrical conductivity (dS cm^{-1}) was determined for 416 out of 563 soil samples that did not already have it according to the methodology employed by the NSSC-KSSL (Soil Survey Staff, 2004).

2.3. Statistical analysis

2.3.1. Mean pH comparisons for soil conditions and measurement methods

An analysis of variance (ANOVA) was used to test the significance of each categorical variable separately (soil order, mineralogy, family particle size classes, genetic master horizons, and depth intervals) on soil pH. The Tukey–Kramer Honestly Significant Difference (HSD) multiple comparison test was conducted to compare the mean soil pH values of each class (Kramer, 1956; Tukey, 1953). The statistical analyses were conducted using JMP Version 9.0 (SAS Institute Inc.; Schlotzhauer, 2007) statistical package. We did not test for interactions between categorical variables on the predicted soil pH.

2.3.2. The influence of soil conditions and measurement methods on predicted pH 1:5_w

The forward step-wise procedure was used in conjunction with adjusted R^2 RMSE and the Akaike Information Criterion (AIC) (Akaike, 1974, 1976) to determine the best predictors of pH 1:5_w. The AIC expression is:

$$\text{AIC} = N \ln R_e + 2p$$

where N is the number of samples, p is the number of parameters in the model being evaluated and R_e is the residual sum of squares. The AIC measures the relative goodness of fit for a model compared to other models with more than one predictor (Akaike, 1974, 1976), and is not necessarily an indicator of the best model. Models with the highest R^2 , lowest RMSE and AIC were selected.

2.3.3. The influence of model technique to the prediction of pH 1:5_w

Simple linear regression analysis was conducted to test the overall significance of the relationship between pH 1:5_w and continuous predictors (pH 1:1_w, pH 1:2_{CaCl2} and EC). Smoothing spline functions (De

Boor, 2001; Green and Silverman, 1994; Reinsch, 1967), artificial neural networks (ANN) (Hastie et al., 2009; Rumelhart and McClelland, 1986), and random forest (RF) (Breiman, 2001) models were also tested. ANN is used for non-linear relationships and mostly for cases where these relationships are not known. Random forest which is a distribution-free method was used because it can better handle a mix of categorical and continuous variables (Díaz-Uriarte and de Andrés, 2006; Izenman, 2008). The RF was performed in R (R Development Core Team, 2011). Finally, the predicted soil pH 1:5_w from all models was regressed against the measured soil pH 1:5_w to examine the model fit using JMP 9.0 (SAS Institute Inc.; Schlotzhauer, 2007).

3. Results and discussions

3.1. Mean pH comparisons by soil conditions and measurement method

The 563 selected soil samples analysed for pH 1:5_w in this study represented a wide range of soil conditions and pH values from 2.9 to 10.5 for a comprehensive evaluation (Table 2). First, we examined the differences in the overall means of soil pH 1:5_w, pH 1:1_w and pH 1:2_{CaCl2} (Table 3). Then we examined the mean differences within soil conditions (soil order, mineralogy classes, family particle size classes, genetic master horizons and depth intervals).

The overall mean value of soil pH 1:2_{CaCl2} across all samples was 5.76, which was significantly lower (p -value < 0.0001) than the mean soil pH 1:1_w (6.34) and pH 1:5_w (6.23). The overall mean of soil pH 1:5_w was significantly lower than the overall mean of pH 1:1_w (p -value < 0.0001), by 0.11 pH units. The lower mean soil pH 1:2_{CaCl2} compared to the mean soil pH 1:1_w and pH 1:5_w has been attributed

Table 2
Summary statistics by methods for soil pH measurement.

Method	Mean	St. dev.	Min.	Median	Max.
pH 1:1 _w	6.2	1.5	3.5	6.0	10.4
pH 1:5 _w	6.3	1.3	3.4	6.2	10.5
pH 1:2 _{Ca}	5.8	1.4	2.9	5.6	10.5

Table 3

Analysis of variance (ANOVA) for the significance of classes and/or grouped mean pH values within each soil condition.

Soil conditions	pH 1:5 _w		pH 1:1 _w		pH 1:2 _{CaCl₂}	
	F ratio	Prob > F	F ratio	Prob > F	F ratio	Prob > F
Soil order	37.88	<0.0001	38.89	<0.0001	32.07	<0.0001
Mineralogy class	28.03	<0.0001	23.21	<0.0001	22.80	<0.0001
Family particle size class	6.68	<0.0001	3.84	0.0041	4.98	0.0006
Genetic master horizon	0.61	0.6000	0.94	0.4198	0.49	0.6825
Depth intervals	0.63	0.7000	0.81	0.5600	0.92	0.4747

to the ionic strength of the soil solution (Aitken and Moody, 1991; Conyers and Davey, 1988; Miller and Kissel, 2010). Addition of salts to the solution lowers the pH due to the exchange of Ca^{2+} with H^+ and Al^{3+} on solid surfaces (Kissel et al., 2009; Miller and Kissel, 2010; Pierre et al., 1970; Schofield and Taylor, 1955). The lower mean of soil pH 1:5_w compared to the mean of pH 1:1_w has been attributed to a dilution effect (Aitken and Moody, 1991; Conyers and Davey, 1988; Miller and Kissel, 2010). For example, Keaton (1938) and Davis (1943) found that increasing the water:soil ratio from 1:10 to 10:1 resulted in a decrease of 0.40 pH units.

For all methods of soil pH determination, there were significant differences between the means of the soil order, mineralogy, and family particle size classes; however, there were no significant differences between the means of the genetic master horizons and depth intervals (Table 3). The lack of difference between depths and genetic master horizons is likely due to the variability of parent materials and other factors with depth across the dataset. Bruce et al. (1989) found that the pH 1:5_w had the same modal range (5.0–5.5) for 91 surface and subsurface samples from acid soil in Queensland, Australia.

Multiple-comparisons of mean pH values (results not shown) show that the mean soil pH of Aridisols was significantly higher than the mean soil pH of the other soil orders for all methods of soil pH determination. Ultisols (4.61–5.27), Spodosols (4.27–4.86), and Histosols (4.83–5.69) had the lowest soil pH. The higher mean value of Aridisols for all pH measurement methods compared to other soil orders is likely due to the accumulation of water-soluble minerals. In environments, with a general lack of precipitation where Aridisols are typically found, these more soluble minerals do not leach out of the profile, with compounds such as calcium carbonate driving higher pH values (Buol et al., 2003; Soil Survey Staff, 2010). Ultisols generally undergo intense weathering and leaching of primary minerals that contain calcium, magnesium and potassium. This leads to relatively acidic soil pH values (Buol et al., 2003; Soil Survey Staff, 2010). The subsurface accumulation of humus–aluminium and humus–iron complexes is partly responsible for low pH values of Spodosols (Buol et al., 2003; Soil Survey Staff, 2010), whereas Histosols found in rain-dependent raised bogs can have soil pH values that vary between 3 and 5.5 (Buol et al., 2003; Soil Survey Staff, 2010).

The overall mean soil pH of the smectitic soils was significantly higher than that of the other mineralogy classes for all methods of soil pH determination. The mean pH values were 7.19 for pH 1:2_{CaCl₂} to 7.81 for pH 1:5_w. The overall mean soil pH of clayey soils was significantly higher than the overall mean soil pH of soils in the other family particle size classes, for all methods of soil pH determination. The mean pH values were 6.29 for pH 1:2_{CaCl₂} and 6.91 for pH 1:5_w. The higher pH values for Vertisols have been linked to their predominately smectitic mineralogy and finer soil texture (Chan et al., 1988; Dang et al., 1994a,b; Uehara and Keng, 1974). For example, Dang et al. (1994a,b) found that the mean pH 1:5 W for 14 Vertisols soil samples in Queensland, Australia was 8.5 (7.5–9.0) with a mean smectitic clay content of 75% (57–91%). Chan et al. (1988) also found the mean pH 1:5 W for 9 Vertisols soil samples to be 7.8 (6.7–8.8) which is higher compared to other soil orders. For this study, 42% of the smectitic samples belonged to Vertisols, which had also the highest pH values for all

pH measurement methods. Also 50% of clayey soils samples were smectitic which explains partly the higher pH values compared to other family particle size classes. The differences between mineralogy classes could not be explained completely by soil order or in the case of family particle size class by mineralogy differences. In this study, it was difficult to separate these effects.

3.2. The influence of model technique on the prediction of pH 1:5_w

3.2.1. Linear regression models

Simple linear regression of the pH 1:5_w on pH 1:1_w and pH 1:2_{CaCl₂} revealed highly significant linear relationships (Table 4). The slope of the regression lines for pH 1:5_w–pH 1:1_w was 1.06 with an RMSE of 0.44 pH units, and for pH 1:5_w–pH 1:2_{CaCl₂} was 1.01 with an RMSE of 0.42 pH units. In each case, more than 90% of the variability in soil pH 1:5_w was explained by pH 1:1_w or pH 1:2_{CaCl₂}.

3.2.2. Multiple linear regression models

The AIC along with adjusted R^2 and RMSE determined the statistical model that minimized the loss of information while keeping the number of model parameters (predictors) at minimum. The “best” model resulting from the step-wise forward procedure that included all continuous and categorical variables was the one with pH 1:1_w, pH 1:2_{CaCl₂}, soil order, mineralogy class, family particle size class, and soil depth with adjusted R^2 0.95 and RMSE 0.33. This model was a slight improvement compared to the simple linear regression model with only pH 1:1_w as a predictor with an R^2 of 0.91 and RMSE of 0.44 or 1:2_{CaCl₂} with an R^2 of 0.92 and RMSE of 0.42 (Fig. 2). The addition of EC (Fig. 3a) as a predictor did not improve the prediction at all. The decrease of the AIC by the addition of categorical variables to pH 1:5_w vs. pH 1:1_w and pH 1:5_w pH 1:2_{CaCl₂} models was highly variable. The largest decrease in AIC was observed for mineralogy class (192, pH 1:1_w) followed by soil order (184, pH 1:1_w). However, for the mineralogy class which was the best case scenario the adjusted R^2 increased by only 0.03 while RMSE decreased by only 0.07 pH units.

Overall, our results indicate that categorical variables did not substantially improve predictions over linear regression with other pH measures. Aherm et al. (1995) found similar results when comparing the influence of surface (0–0.1 m), subsurface (0.1–0.2 m; 0.2–0.3 m) and subsoil (>0.3 m) depth categories on the relationship between pH 1:5_w and pH_{CaCl₂}. The distribution of the residual errors between

Table 4

Model parameter influence on predicted pH 1:5_w from the forward step-wise procedure.

Model parameters	Adjusted R^2	RMSE	P value	AIC
pH 1:1 _w	0.91	0.44	<0.0001	676
pH 1:2 _{CaCl₂}	0.92	0.42	<0.0001	621
pH 1:1 _w & 1:2 _{CaCl₂}	0.93	0.40	<0.0001	570
pH 1:1 _w & soil order	0.94	0.37	<0.0001	489
pH 1:2 _{CaCl₂} & soil order	0.95	0.35	<0.0001	441
pH 1:1 _w & 1:2 _{CaCl₂} & soil order	0.96	0.33	<0.0001	383
pH 1:1 _w & mineralogy class	0.94	0.39	<0.0001	484
pH 1:2 _{CaCl₂} & soil mineralogy	0.94	0.38	<0.0001	473
pH 1:1 _w & 1:2 _{CaCl₂} & mineralogy class	0.95	0.36	<0.0001	433
pH 1:1 _w & family particle size class	0.92	0.43	<0.0001	650
pH 1:2 _{CaCl₂} & family particle size class	0.93	0.41	<0.0001	596
pH 1:1 _w & 1:2 _{CaCl₂} & family particle size class	0.94	0.38	<0.0001	526
pH 1:1 _w & genetic master horizons	0.92	0.44	<0.0001	655
pH 1:2 _{CaCl₂} & genetic master horizons	0.93	0.42	<0.0001	590
pH 1:1 _w & 1:2 _{CaCl₂} & genetic master horizons	0.94	0.38	<0.0001	546
pH 1:1 _w & Soil depth intervals	0.92	0.44	<0.0001	685
pH 1:2 _{CaCl₂} & soil depth intervals	0.92	0.41	<0.0001	620
pH 1:1 _w & 1:2 _{CaCl₂} & soil depth intervals	0.93	0.40	<0.0001	572
pH 1:1 _w & electrical conductivity	0.91	0.44	<0.0001	676
pH 1:2 _{CaCl₂} & electrical conductivity	0.92	0.42	<0.0001	621
pH 1:2 _w & 1:2 _{CaCl₂} & electrical conductivity	0.93	0.40	<0.0001	575

RMSE Root Mean Square Error.

AIC Akaike Information Criterion.

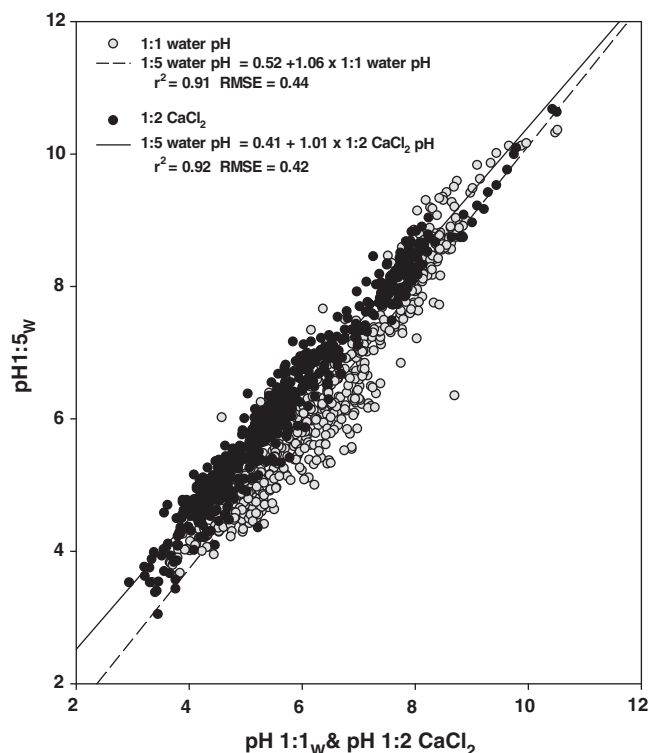


Fig. 2. Linear regression model of pH 1:5 water (pH 1:5_W) versus pH 1:1 water (pH 1:1_W) (black circles) and pH 1:5_W versus pH 1:1_W (open circles).

classes within categorical variables was not over or under predicting the pH 1:5_W (results not shown). For both measured pH 1:1_W and pH 1:2_{CaCl₂} predictors, Aridisols and Vertisols had the highest residuals among soil orders while for family particle size classes the clayey class had the highest residuals.

3.2.3. Other models

3.2.3.1. Smoothing spline models. In our study, we applied the smoothing spline function to the pH 1:5_W vs. pH 1:2_{CaCl₂} and pH 1:5_W vs. pH 1:1_W. For both fits, the adjusted R^2 was 0.93—an increase of only 0.01 over the linear fit of pH 1:5_W–pH 1:2_{CaCl₂}, and an increase of 0.02 over the linear fit of pH 1:5_W–pH 1:1_W. The smoothing spline resulted in a slight decrease in RMSE from 0.42 for the linear model to 0.40 (pH 1:5_W vs. pH 1:2_{CaCl₂}) and from 0.44 to 0.41 (pH 1:5_W vs. pH 1:1_W).

3.2.3.2. Artificial neural network and random forest. The ANN and RF models using both pH 1:1_W and pH 1:2_{CaCl₂} EC and categorical variables as predictors were not different from each other and performed similarly compared to the linear models. The RMSE was only 0.37 for ANN using pH 1:1_W, pH 1:2_{CaCl₂}, and EC as predictors. This was slightly less compared to the models using pH 1:1_W and EC or 1:2_{CaCl₂} and EC with RMSE 0.38 and 0.43. The combination of EC and pH 1:1_W as predictors using RF resulted in an RMSE = 0.39. The addition of pH 1:2_{CaCl₂} decreased the RMSE to 0.33. Because the distribution of EC was skewed as found by other researchers (Minasny et al., 2011) the models with log transformed EC were also tested. The models with log transformed and un-transformed EC were not different from each other with respect to adjusted R^2 and RMSE.

3.2.4. Model selection

The analysis of our limited data set (563 soil samples) indicates that the simple linear model is adequate for predicting pH 1:5_W and along with AIC provides further support for the selection of the simple linear regression model versus other models. More complex models did not substantially improve the R^2 and RMSE which makes the linear model a more suitable candidate for use in conversion of soil pH data for GlobalSoilMap.

In order to support the selection of simple models, the slight improvements in RMSE and R^2 for more complex models need to be placed in the context of other sources of errors. For example, the comparison of the errors associated with the analytical methods for soil pH

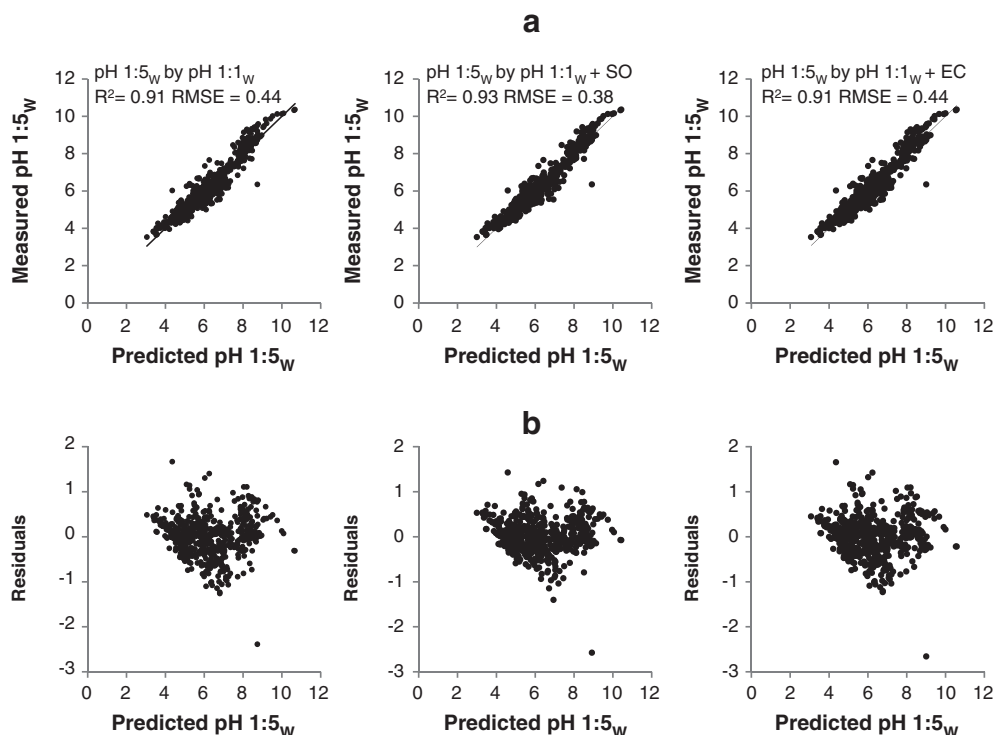


Fig. 3. (a) Linear regression model and (b) residuals of measured pH 1:5_W versus predicted pH 1:5_W by models using pH 1:1_W, soil order (SO) and electrical conductivity (EC) predictors.

Table 5

Error source and magnitude associated with pH 1:1_w and pH 1:2_{CaCl₂} predictors and analytical pH methods.

Error source	No.	RMSE	
		obs.	pH 1:1 _w pH 1:2 _{CaCl₂}
Simple linear regression models for predicting pH 1:5 _w	563	0.4	0.4
Between laboratory reproducibility (<i>R</i>)	16	0.7	0.5
Within-laboratory reproducibility (<i>R_L</i>)	183	0.3	0.2
Repeatability (<i>r</i>)	199	0.2	0.1

measurements and RMSE from using continuous (pH 1:1_w, pH 1:2_{CaCl₂}, and EC) and/or categorical predictors (soil order, mineralogy classes and family particle size classes) indicates that analytical methods contribute similar uncertainties to the predicted soil pH 1:5_w (Table 5).

As part of the International Soil Exchange the NSSC-NSSL receives soil samples quarterly from Wageningen University (Wageningen, 2010) for analysis by a suite of methods, including pH 1:1_w and pH 1:2_{CaCl₂}. NSSC-KSSL reports three different estimates of precision for routine pH measurements: (i) between-laboratory reproducibility (*R*), (ii) within-laboratory reproducibility (*R_L*) and (iii) repeatability (*r*). Each of these precision estimates measures the spread of results when a sample is analysed repetitively under specified conditions, with 95% confidence that results would fall within \pm (*R*, *R_L*, or *r*) of the mean.

In our study, the RMSE for the simple linear regression model using only pH 1:1_w was 0.44 pH units, and for the model using only pH 1:2_{CaCl₂} it was 0.42 pH units (Table 3). These are comparable with the error estimates from analytical methods (Table 5).

The use of smoothing spline functions, ANN, or RF did not significantly improve the predictions for our data. This is possibly due to the limited number of soil samples in our calibration dataset and the poorly-expressed dilution effect, especially for pH 1:1_w vs. pH 1:5_w. In general, previous studies have found that the relationship between pH dilutions becomes more obviously non-linear as sample size increases. Aitken and Moody (1991) were among the first to question the linearity of the relationship but their sample size of 90 samples was not large enough to draw reasonable conclusions. Ahern et al. (1995) and Henderson and Bui (2002) used a larger number of samples to provide evidence for the lack of linearity of the relationship between pH methods. Our sample size was approximately 2 orders of magnitude smaller than that of Henderson and Bui (2002) and one order of magnitude less than Ahern et al. (1995), we observed non-linearity as well. We argue that as the sample size increases the scarcity of samples in the extreme pH ranges is enhanced reflecting in part the natural abundance of such extreme pH values and/or the bias in sample collection.

Our results and those from other research (Aitken and Moody, 1991; Miller and Kissel, 2010) indicate that the relationship between pH dilutions is largely linear with a slope of close to 1. However, some authors have found a non-linear relationship between pH 1:5_w and pH 1:5_{CaCl₂} across the entire range of soil pH (Ahern et al., 1995; Aitken and Moody, 1991; Bruce et al., 1989; Henderson and Bui, 2002). In our study, the slope of the regression line for pH 1:5_w vs. pH 1:2_{CaCl₂} was 1.01 indicating that the difference between pH 1:5_w and pH 1:2_{CaCl₂} increases slightly as soil pH increases. Little (1992) suggested that, in the alkaline pH region, buffering by calcium carbonate is responsible for the increasing difference, while in extremely acid regions, the displacement and hydrolysis of Al³⁺ from exchange sites by CaCl₂ buffer the system. Aitken and Moody (1991) suggested the variable charge theory as an explanation for the tendency of the difference between pH 1:5_w and pH 1:5_{CaCl₂} values to decrease with decreasing pH values. They fitted quadratic functions and increased the coefficient of determination from 0.88 (linear) to 0.92 (quadratic model) to account for this. Ahern et al. (1995) found the linear fit suitable only for a restricted range of soil pH values (4.5–9.0). They also offered a quadratic fit (second-order polynomial) but declared that even the quadratic fit was not suitable for a wider range of soil pH values. Little (1992) found a cubic fit

(third-order polynomial) maximised the *R*² and was a better representation of the relationship at the extreme ranges of pH values. Henderson and Bui (2002) applied a smoothing spline function with six degrees of freedom for converting the soil pH 1:5_w of 70,465 samples to pH 1:5_{CaCl₂}. The samples had a range of pH between 2.5 and 10.5. They found the smoothing spline model with an *R*² of 0.96 to be stable for values outside of the range of the data and to significantly reduce the residuals compared to the linear models. In our study, the range in soil pH was 3.5–10.5 across all methods. Minasny et al. (2011) found that the difference between pH 1:5_w and pH 1:5_{CaCl₂} decreased with increasing soil EC. Using ANN, which had a curvilinear relationship between pH 1:5_w and pH 1:5_{CaCl₂}, Minasny et al. (2011) offered better predictions for acid and alkaline soil pH; however, both linear and ANN models performed similarly. The ANN in our study performed similarly with the linear models confirming the findings from Minasny et al. (2011).

We argue that simplicity of the linear model using either pH 1:1_w or pH 1:2_{CaCl₂} as the only predictor outweighs the benefits of adding more, categorical predictors (soil order, mineralogy or family particle size classes) or fitting more complex models, especially given the fact that the laboratory analytical errors could be as large as the RMSE of various models. Furthermore, this simple equation enables the prediction of pH 1:5_w where information on soil conditions (taxonomy, horizon, texture or mineralogy) is not always available.

4. Conclusions

The wide range of soil sample conditions represented in our study allowed for a comparison of the soil pH measurements between soil orders, mineralogy classes, family textural classes, genetic master horizons and depth intervals. Despite significant differences between the within-class soil pH means of a number of soil order, mineralogy and family particle size classes, they explained a small amount of soil pH 1:5_w variability and did not improve predictions. We also found the addition of EC combined with the use of smoothing spline function, ANN and RF did not substantially improve the prediction of pH 1:5_w.

We conclude that simple linear regression models using only pH 1:1_w or pH 1:2_{CaCl₂} are suitable for predicting pH 1:5_w. It is possible that a more extensive set of samples would support the fitting of the data with other models. The errors associated with these predictive models are comparable to those associated with the analytical methods alone. Furthermore, the simplicity of the linear model with pH 1:1_w or pH 1:2_{CaCl₂} as the only predictor outweighed the benefits of adding more categorical parameters (soil order, mineralogy classes or family particle size classes) or continuous ones (EC) to the model. The additional parameters did not substantially improve the prediction and may not always be available for making predictions. Our results indicate that these simple linear regression equations are adequate to generate a predictive soil pH 1:5_w map for standard depths (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm) for GlobalSoilMap.

A total of 563 soil samples representing 98 pedons from the SCDB and physically archived at the KSSL that had been previously analyzed for soil pH 1:1_w and pH 1:2_{CaCl₂} were selected for determination of soil pH 1:5_w, pH 1:5_{CaCl₂}, and EC. In this paper, only the results for the conversion of soil pH 1:1_w and pH 1:2_{CaCl₂} to pH 1:5_w along with the influence of EC will be presented (data can be downloaded at: <ftp://ftp-fc.sc.egov.usda.gov/NSSC/pub/geoderma/>, contact webmaster@lin.usda.gov for any downloading issues).

Acknowledgements

The authors would like to thank Patty Jones and Larry Arnold, the Lead Physical Science Technicians at the NRCS-NSSC, and Heather Hansen for their help with the additional soil analysis of soil samples and technical editing efforts. We also would like to extend our thanks to the reviewers whose comments and suggestions significantly improved this paper.

References

- Ahern, C.R., Baker, D.E., Aitken, R.L., 1995. Models for relating pH measurements in water and calcium chloride for a wide range of pH, soil types and depths. *Plant Soil* 171, 47–52.
- Aitken, R.L., Moody, P.W., 1991. Interrelations between soil pH measurements in various electrolytes and soil solution pH in acidic soils. *Aust. J. Soil Res.* 29, 483–491.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19 (6), 716–723. <http://dx.doi.org/10.1109/TAC.1974.1100705>.
- Akaike, H., 1976. An information criterion (AIC). *Math. Sci.* 14 (153), 5–9.
- Brady, N.C., Weil, R.R., 1999. *The Nature and Properties of Soil*, 12th edition. Prentice-Hall, Inc., New Jersey.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Bruce, R.C., Warrell, L.A., Bell, L.C., Edwards, D.G., 1989. Chemical attributes of some Queensland acid soils. I. Solid and solution phase compositions. *Aust. J. Soil Res.* 27, 333–351.
- Buol, S.W., Southard, R.J., Graham, R.C., McDaniel, P.A., 2003. *Soil Genesis and Classification*, 5th edition. Iowa State Press, Ames, Iowa.
- Chan, K.Y., Bellotti, W.D., Roberts, W.P., 1988. Changes in surface soil properties of Vertisols under dryland cropping in a semi-arid environment. *Aust. J. Soil Res.* 26, 509–518.
- Conyers, M.K., Davey, B.G., 1988. Observations on some routine methods for soil pH determination. *Soil Sci.* 145, 29–36.
- Dang, Y.P., Dalal, R.C., Edwards, D.G., Tiller, K.G., 1994a. Kinetics of zinc desorption from Vertisols. *Soil Sci. Soc. Am. J.* 58, 1392–1399.
- Dang, Y.P., Tiller, K.G., Dalal, R.C., Edwards, D.G., 1994b. Zinc speciation in soil solution of Vertisols. *Aust. J. Soil Res.* 34, 369–383.
- Davis, L.E., 1943. Measurements of pH with the glass electrode as affected by soil moisture. *Soil Sci.* 56, 405–422.
- De Boor, C., 2001. *A Practical Guide to Splines*, Revised edition. Springer.
- R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Díaz-Uriarte, R., de Andrés, S.A., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 3 (7). <http://dx.doi.org/10.1186/1471-2105-7-3>.
- GlobalSoilMap, 2012. Specifications, Version 1 GlobalSoilMap.Net Products. Release 2.2.
- Green, P.J., Silverman, B.W., 1994. *Nonparametric Regression and Generalized Linear Models*. CRC Press.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Series in Statistics, 2nd ed. Springer-Verlag, New York.
- Henderson, B.L., Bui, E.N., 2002. An improved calibration curve between soil pH measured in water and CaCl_2 . *Aust. J. Soil Res.* 40, 1399–1405.
- Izenman, A.J., 2008. *Modern Multivariate Statistical Techniques Regression, Classification and Manifold Learning*. Springer Series in Statistics, New York.
- Keaton, C.M., 1938. A theory explaining the relation of soil–water ratios to pH values. *Soil Sci.* 46, 259–265.
- Kissel, D.E., Sonon, L., Vendrell, P.F., Isaac, R.A., 2009. Salt concentration and measurement of soil pH. *Commun. Soil Sci. Plant Anal.* 40, 179–187.
- Kramer, C.Y., 1956. Extension of multiple range tests to group means with unequal number of replications. *Biometrics* 12, 307–310.
- Little, I.P., 1992. The relationship between soil pH measurements in calcium chloride and water suspensions. *Aust. J. Soil Res.* 30, 587–592.
- Miller, R.O., Kissel, D.E., 2010. Comparison of soil pH methods on soils of North America. *Soil Sci. Soc. Am. J.* 74, 310–316.
- Minasny, B., McBratney, A.B., Brough, D.M., Jacquier, D., 2011. Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. *Eur. J. Soil Sci.* <http://dx.doi.org/10.1111/j.1365-2389.2011.01386.x>.
- National Cooperative Soil Survey, 2011. National Cooperative Soil Characterization Database. Available online at <http://ncsslabdatamart.sc.egov.usda.gov> (Accessed [July 2011]).
- Odgers, N.P., Libohova, Z., Thompson, J.A., 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. *Geoderma* 189–190, 153–163. <http://dx.doi.org/10.1016/j.geoderma.2012.05.026>.
- Pierre, W.H., Meisinger, J., Birchett, J.R., 1970. Cation–anion balance in crops as a factor in determining the effect of nitrogen fertilizer on soil acidity. *Agron. J.* 62, 106–112.
- Rayment, G.E., Higginson, F.R., 1992. *Australian laboratory handbook of soil and water chemical methods*. Australian Soil and Land Survey Handbook, vol. 3. Inkata Press, Melbourne.
- Reinsch, Ch.H., 1967. Smoothing by spline functions. *Numer. Math.* 10, 177–183.
- Rumelhart, D., McClelland, J., 1986. *Parallel Distributed Processing*. MIT Press, Cambridge, Mass.
- Schlotzhauer, S.D., 2007. *Elementary Statistics Using JMP*. SAS Institute Inc., Cary, NC.
- Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., Soil Survey Staff, 2012. *Field Book for Describing and Sampling Soils*, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.
- Schofield, R.K., Taylor, A.W., 1955. The measurement of soil pH. *Soil Sci. Soc. Am. Proc.* 19, 164–167.
- Soil Survey Staff, 1999. *Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys*, 2nd edition.
- Soil Survey Staff, 2004. *Soil Survey Laboratory Methods Manual*. In: Burt, R. (Ed.), *Soil Survey Investigations Report No. 42*, Version 4.0. U.S. Department of Agriculture, Natural Resources Conservation Service.
- Soil Survey Staff, 2009. *Soil Survey Field and Laboratory Methods Manual*. In: Burt, R. (Ed.), *Soil Survey Investigations Report No. 51*, Version 1.0. U.S. Department of Agriculture, Natural Resources Conservation Service.
- Soil Survey Staff, 2010. *Keys to Soil Taxonomy*, 11th edition. U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, D.C.
- Soil Survey Staff, 2011a. Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. (<http://sdmdataaccess.nrcs.usda.gov/>. [Accessed Dec 2011]).
- Soil Survey Staff, 2011b. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. (Available online at <http://websoilsurvey.nrcs.usda.gov/>. [Accessed Aug 2011]).
- Tukey, J.W., 1953. Some selected quick and easy methods of statistical analysis. *Trans. N. Y. Acad. Sci.* 16 (2), 88–97.
- Uehara, G., Keng, J.Ch., 1974. Management implications of soil mineralogy in Latin America. Presented at Symposium on Management of Tropical Soils, Cali, Columbia (21 pp.).
- United States Department of Agriculture, Natural Resources Conservation Service, 2011. National Cooperative Soil Survey, National Cooperative Soil Characterization Database. Available online at <http://ncsslabdatamart.sc.egov.usda.gov> (Accessed on January 2011).
- Wagenin, 2010. Evaluating Programs for Analytical Laboratories International Soil Analytical Exchange Quarterly Reports. 1–4.